



Resample-smoothing of Voronoi intensity estimators

M. Mehdi Moradi¹ · Ottmar Cronie² · Ege Rubak³ · Raphael Lachize-Rey⁴ · Jorge Mateu⁵ · Adrian Baddeley⁶

Received: 28 June 2018 / Accepted: 21 December 2018
© The Author(s) 2019

Abstract

Voronoi estimators are non-parametric and adaptive estimators of the intensity of a point process. The intensity estimate at a given location is equal to the reciprocal of the size of the Voronoi/Dirichlet cell containing that location. Their major drawback is that they tend to paradoxically under-smooth the data in regions where the point density of the observed point pattern is high, and over-smooth where the point density is low. To remedy this behaviour, we propose to apply an additional smoothing operation to the Voronoi estimator, based on resampling the point pattern by independent random thinning. Through a simulation study we show that our resample-smoothing technique improves the estimation substantially. In addition, we study statistical properties such as unbiasedness and variance, and propose a rule-of-thumb and a data-driven cross-validation approach to choose the amount of smoothing to apply. Finally we apply our proposed intensity estimation scheme to two datasets: locations of pine saplings (planar point pattern) and motor vehicle traffic accidents (linear network point pattern).

Keywords Adaptive intensity estimation · Independent thinning · Machine learning · Point process · Resampling · Voronoi intensity estimator

1 Introduction

In the analysis of spatial point patterns (van Lieshout 2000; Chiu et al. 2013; Diggle 2014; Baddeley et al. 2015), exploratory investigation often starts with non-parametric analysis of the spatial intensity of points. The *intensity function*, which is a first order moment characterisation of the

point process assumed to have generated the data, reflects the abundance of points in different regions and may be seen as a “heat map” for the events. For most datasets, it is not realistic to assume that the underlying point process is homogeneous, i.e. that its intensity function is constant; rather it is natural to start by assuming inhomogeneity.

The most prominent approach to non-parametric intensity estimation is undoubtedly kernel estimation (Diggle 1985; Silverman 1986; Diggle 2014; Baddeley et al. 2015). The degree of smoothing is controlled by a smoothing parameter, called the bandwidth, and the resulting estimates heavily depend on the choice of bandwidth. A small bandwidth may result in under-smoothing whereas a large bandwidth might result in over-smoothing the intensity. Data-based procedures for bandwidth selection have been studied extensively (Diggle 1985; Silverman 1986; Berman and Diggle 1989; Scott 1992; Wand and Jones 1995; Loader 1999) including some recent advances (Cronie and van Lieshout 2018). A further problem with kernel estimation is that, if there are wide variations in intensity across the spatial domain, it may be impossible to find a single fixed bandwidth value which is satisfactory for smoothing every part of the spatial domain. Consequently the bandwidth must be spatially-varying, giving rise to a spatially “adaptive” kernel estimator (Davies and

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11222-018-09850-0>) contains supplementary material, which is available to authorized users.

✉ Ottmar Cronie
ottmar.cronie@umu.se

- ¹ Institute of New Imaging Technologies (INIT), University Jaume I, Castellón, Spain
- ² Department of Mathematics and Mathematical Statistics, Umeå University, Umeå, Sweden
- ³ Department of Mathematical Sciences, Aalborg University, Aalborg, Denmark
- ⁴ Université Paris Descartes, Sorbonne Paris Cité, Paris, France
- ⁵ Department of Mathematics, University Jaume I, Castellón, Spain
- ⁶ Department of Mathematics and Statistics, Curtin University, Perth, Australia

Hazelton 2010; Diggle 2014; Davies et al. 2016; Davies and Baddeley 2018) at the cost of increased complexity.

Recently there has been increasing interest in point patterns on linear networks (Okabe and Sugihara 2012; Ang et al. 2012; Baddeley et al. 2015; Rakshit et al. 2018); examples include street crimes or traffic accidents on a road network (of a city). Here the matter of kernel estimation is even more delicate due to the geometry of the underlying network. Borruo (2003, 2005, 2008) proposed several methods for kernel smoothing of network data without discussing statistical properties. Xie and Yan (2008) defined a kernel-based intensity estimator for network point patterns without taking the topography of the network into consideration and as a result the estimation errors tended to be large, thus making the estimator heavily biased. Okabe et al. (2009) further introduced a class of so-called equal-split network kernel density estimators which support both continuous and discontinuous schemes. By exploiting properties of diffusion on networks, McSwiggan et al. (2017) developed a kernel estimation method based on the heat kernel, which is the appropriate linear network analogue of the Gaussian kernel. In addition, Moradi et al. (2018) extended the classical spatial edge corrected kernel intensity estimator to point patterns on linear networks.

As a consequence of underlying causes such as demography and human mobility, it is quite common to encounter sharp boundaries between high and low concentrations of events. For example, street crimes and traffic accidents tend to happen particularly in busy streets, which may be surrounded by quiet neighbourhoods. The classical kernel estimation approach is often unsuitable for such types of data.

Echoing Barr and Schoenberg (2010), we argue that kernel-based approaches may be unsatisfactory when there are sharp boundaries between parts with high and low intensities. Fixed bandwidth kernel smoothing results in over-smoothing in high-intensity areas, under-smoothing in low-intensity areas, and a blurring of sharp boundaries (Baddeley et al. 2015). By using a spatially adaptive kernel estimator we may reduce such problems when estimating the intensity function, but optimal bandwidth selection becomes even more challenging and important (Davies and Hazelton 2010).

As an alternative, one could consider an approach without any choice of tuning parameters, e.g. a tessellation-based approach (van Lieshout 2012; Schaap 2007). One such approach is provided by Voronoi intensity estimation (Ord 1978; Barr and Schoenberg 2010; Okabe and Sugihara 2012), defined such that within a given Voronoi cell of the point pattern the intensity estimate is set to the reciprocal of the size of that cell (Okabe et al. 2000). When employing the Voronoi intensity estimator, one thing that quickly becomes evident is that it often accentuates local features too much, in particular in regions with high event density. This reflects a

previously observed phenomenon: adaptive estimators, such as the Voronoi intensity estimator, may smooth too little whereas kernel estimators may smooth too much in dense regions (Baddeley et al. 2015, Section 6.5.2). Hence, one should be able to find some middle ground and we here aim at providing a contribution to that.

Our idea is simple. In dense parts surrounded by empty neighbourhoods, Voronoi intensity estimators tend to smooth too little, thus generating excessive peaks in the intensity estimate in those parts. By removing points in such a dense part we reduce the peaks, which results in a smoother intensity estimate, with a shape more similar to the true intensity function. However, the problem of doing this “manually” is twofold: (1) it is not clear which specific points we should remove, and (2) we need to compensate for the reduced total mass. To solve these issues, we propose to generate $m \geq 1$ independent random point patterns, each obtained by randomly thinning the original point pattern with the same retention probability p . From each of the thinned patterns we compute a Voronoi intensity estimate. In order to compensate for the reduced mass, we then scale each of the m estimates by the reciprocal of the retention probability, and use the corresponding average as final estimate of the intensity function. We propose this technique for point patterns in rather general spaces.

The paper is structured as follows. In Sect. 2 we give a short background on point processes and intensity estimation. In Sect. 3 we introduce our resample-smoothing technique, study its statistical properties and discuss ways to choose the amount of smoothing, i.e. thinning, to apply. In Sect. 4 we assess the performance of our approach numerically for a few different planar point processes and in Sect. 5 we apply our methodology to two datasets: a planar point pattern and a linear network point pattern. Section 6 contains a discussion and some directions for future work and in the Electronic Supplementary Material we provide the proofs of the theoretical results in the paper as well as bias and variance plots together with box plots for estimation errors for the simulation study in Sect. 4.

2 Preliminaries

The spatial domain is a general space S , assumed to be a complete separable metric space with distance metric $d(\cdot, \cdot)$. Assume there is a reference measure $A \mapsto |A|$ for $A \subseteq S$, which is sigma-finite and locally finite. Integration with respect to this measure is denoted by $\int du$. All subsets $A \subseteq S$ under consideration are assumed to be Borel sets.

Let X be a simple point process (Daley and Vere-Jones 2008) in S . A realisation of X is a locally-finite set of points in S . The cardinality of the set $X \cap A$, $A \subseteq S$, will be denoted by $N(X \cap A) \in \{0, 1, \dots\}$ and we note that by definition we have

$N(X \cap A) < \infty$ a.s. for bounded $A \subseteq S$ and $N(X \cap \{u\}) \in \{0, 1\}$ for any $u \in S$.

A *point pattern* is a finite set $\mathbf{x} = \{x_1, \dots, x_n\} \subset S$, $n \geq 0$, of distinct points in S . Inside any bounded study region $W \subseteq S$, the partial realisation $X \cap W$ of the point process is a point pattern.

Relevant examples of the space S include:

- Euclidean space $S = \mathbb{R}^d$ of dimension $d \geq 1$ (van Lieshout 2000; Diggle 2014; Baddeley et al. 2015) with the Euclidean distance $d(u, v) = \|u - v\|$, $u, v \in \mathbb{R}^d$, where $\|\cdot\| = \|\cdot\|_d$ denotes the Euclidean norm, and Lebesgue measure $|\cdot|$.
- The sphere $S = \alpha S^{d-1} = \{x \in \mathbb{R}^d : \|x\|_d = \alpha\}$, of radius $\alpha > 0$ in dimension $d \geq 1$, where $d(\cdot, \cdot)$ is the great circle distance and $|\cdot|$ is the spherical surface measure (Lawrence et al. 2016; Møller and Rubak 2016).
- A linear network, i.e. a union

$$S = L = \bigcup_{i=1}^k l_i$$

of $k \in \{1, 2, \dots\}$ line segments $l_i = [u_i, v_i] = \{tu_i + (1-t)v_i : 0 \leq t \leq 1\} \subseteq \mathbb{R}^d$, $d \geq 1$. A common choice for $d(u, v)$ is the shortest-path distance, which gives the shortest length of any path in L joining $u, v \in L$ (Okabe and Sugihara 2012; Ang et al. 2012; Rakshit et al. 2017). Treated as a graph with vertices given by the intersections and endpoints of the line segments, L is assumed to be connected. The measure $|\cdot|$ here corresponds to integration with respect to arc length.

We emphasise that in each of the above cases there exist other metrics and measures which may be more suited for a particular context (Rakshit et al. 2017).

At times, we will assume that X is stationary, or invariant. More specifically, there is a family of transformations/shifts $\{\theta_s : s \in S\}$, $\theta_s : S \rightarrow S$, along S , which induces a so-called flow, under which the distribution of $\theta_s X = \{\theta_s(x) : x \in X\}$ coincides with that of X for any $s \in S$. The underlying assumption will be that S is a so-called (unimodular) homogeneous space with a fixed origin $o \in S$, with $d(\cdot, \cdot)$ chosen such that it metrizes S and $|\cdot|$ chosen to be the associated (left) Haar measure (Last 2010; Schneider and Weil 2008); each such space is a locally compact second-countable Hausdorff space and thereby S becomes a complete separable metric space. To exemplify, in Euclidean spaces with $|\cdot|$ chosen to be Lebesgue measure, we let $\theta_s(u) = u + s \in \mathbb{R}^d$, $u, s \in \mathbb{R}^d$, which yields the classical notion of stationarity, and on a sphere with the corresponding spherical measure we consider the orthogonal group of rotations. Note that a more general setting is also possible (Kallenberg 2017, Chapter 7).

2.1 Intensity functions

To characterise the first moment of X , i.e. the marginal distributional properties of its points, we consider its intensity function $\rho : S \rightarrow [0, \infty)$. It may be defined through the Campbell formula (Daley and Vere-Jones 2008) which states that for any measurable function $f \geq 0$ on S ,

$$\mathbb{E} \left[\sum_{x \in X} f(x) \right] = \int_S f(u) \rho(u) du.$$

In particular,

$$\mathbb{E}[N(X \cap A)] = \int_A \rho(u) du$$

for any $A \subseteq S$. If X is stationary, then $\rho(u) \equiv \rho \in (0, \infty)$ for any $u \in S$. Heuristically, $\rho(u) du$ may be interpreted as the probability of finding a point of X in an infinitesimal neighbourhood du of $u \in S$ with measure du .

2.2 Independent thinning

A key ingredient in our smoothing technique is the notion of independent thinning (Chiu et al. 2013, Section 5.1): given some measurable retention probability function $p(u) \in (0, 1]$, $u \in S$, we run through the points of X and delete a point $x \in X$ with probability $1 - p(x)$, independently of the deletions carried out for the other points of X . The resulting thinned process has intensity

$$\rho_{th}(u) = p(u) \rho(u), \quad u \in S,$$

where $\rho(\cdot)$ is the intensity of the original process X (Chiu et al. 2013, Section 5.1). For further details on the thinning of point processes, see e.g. Møller and Schoenberg (2010) and Daley and Vere-Jones (2008, Section 11.3).

It is worth mentioning that a Poisson process stays Poissonian after independent thinning (Daley and Vere-Jones 2008, Exercise 11.3.1) and, in addition, the independent thinning of an arbitrary point process X with low retention probability results in a point process which, from a distributional point of view, is approximately a Poisson process (Baddeley et al. 2015, Section 9.2.2).

2.3 Voronoi tessellations

The next key ingredient in our estimation scheme is the Voronoi/Dirichlet tessellation of a point pattern $\mathbf{x} = \{x_1, \dots, x_n\}$ contained in some subset $W \subseteq S$ (Chiu et al. 2013; Okabe et al. 2000). Generally speaking, a tessellation of W is a tiling such that i) the union of all tiles constitutes all of W , and ii) the interiors of any two tiles have empty intersections.

The Voronoi/Dirichlet cell \mathcal{V}_x associated with $x \in \mathbf{x}$ consists of all $u \in S$ which are closer to x than any $y \in \mathbf{x} \setminus \{x\}$, i.e.

$$\mathcal{V}_x = \mathcal{V}_x(\mathbf{x}, W) = \{u \in W : d(x, u) \leq d(y, u) \text{ for all } y \in \mathbf{x} \setminus \{x\}\}. \quad (1)$$

The tiling $\{\mathcal{V}_x\}_{x \in \mathbf{x}}$ is termed the Voronoi/Dirichlet tessellation generated by \mathbf{x} . Clearly, the shape of each \mathcal{V}_x depends on the distance $d(\cdot, \cdot)$ chosen for S and its size, $|\mathcal{V}_x|$, depends on the chosen reference measure $|\cdot|$.

2.4 Intensity estimation

Given a point pattern $\mathbf{x} = \{x_1, \dots, x_n\}$ in some study region $W \subseteq S$, $|W| > 0$, we next set out to estimate $\rho(u)$, $u \in W$, under the assumption that \mathbf{x} is a realisation of $X \cap W$.

Before going into details about specific estimators, we briefly mention how different estimators' performances may be evaluated and compared. To evaluate the performance of an estimator $\hat{\rho}(\cdot) = \hat{\rho}(\cdot; X, W)$ of $\rho(u)$, $u \in W$, it is common practice to employ the *Mean Integrated Square Error* (MISE):

$$\begin{aligned} \text{MISE} &= \mathbb{E} \left[\int_W (\hat{\rho}(u) - \rho(u))^2 du \right] \\ &= \int_W \text{Var}(\hat{\rho}(u)) du + \int_W \text{bias}(\hat{\rho}(u))^2 du \\ &= \text{IV} + \text{ISB}, \end{aligned} \quad (2)$$

where $\text{bias}(\hat{\rho}(u)) = \mathbb{E}[\hat{\rho}(u)] - \rho(u)$. Given $k \geq 1$ realisations of $X \cap W$, to obtain an estimate of MISE we average over the integrated square errors generated by each of the k realisations.

Alternatively, we may find estimates of the functions $\text{Var}(\hat{\rho}(u))$ and $\text{bias}(\hat{\rho}(u))$, $u \in W$, based on the k patterns and integrate these over W . This is the setup chosen for the numerical evaluations presented in Sect. 4.

2.4.1 Voronoi intensity estimation

In practice, it is often the case that events occur frequently in specific parts of the study region, e.g. that accidents often happen in more crowded streets or on specific parts of a highway, or that trees tend to grow mainly in specific parts of a forest. In other words, there are sharp boundaries between parts with high and low intensities. We argue, similarly to Barr and Schoenberg (2010) and Ogata (2011), that in order not to blur such boundaries, it is preferable to employ an adaptive intensity estimation scheme, which adapts locally to changes in the spatial distribution of the events.

Here we focus on a particular kind of adaptive intensity estimator, the Voronoi estimator, defined as follows.

Definition 1 For a point process X with intensity function $\rho(\cdot)$, the Voronoi intensity estimator of $\rho(u)$, $u \in W \subseteq S$, $|W| > 0$, is given by

$$\begin{aligned} \hat{\rho}^V(u) &= \hat{\rho}^V(u; X, W) = \sum_{x \in X \cap W} \frac{\mathbf{1}\{u \in \mathcal{V}_x\}}{|\mathcal{V}_x|} \\ &= \sum_{x \in X \cap W} \frac{\mathbf{1}\{u \in \mathcal{V}_x(X, W)\}}{|\mathcal{V}_x(X, W)|}, \quad u \in W, \end{aligned} \quad (3)$$

where \mathcal{V}_x is the Voronoi cell defined in (1). If $X \cap W = \emptyset$ then $\hat{\rho}^V(u) = 0$.

The Voronoi intensity estimator, which was introduced by Brown (1965) and Ord (1978) in the context of Euclidean spaces, has been considered by Baddeley (2007); Ogata (2011); Barr and Schoenberg (2010); van Lieshout (2012). Ebeling and Wiedenmann (1993) have used it to study local spatial concentration of photons, Duyckaerts et al. (1994) and Duyckaerts and Godefroy (2000) have employed it to estimate neuronal density, and it has been applied in the setting of statistical seismology by Ogata (2011) and Baddeley et al. (2015). In the context of linear networks, Okabe and Sugihara (2012) discussed a Voronoi based density estimator, the network Voronoi cell histogram, for the purpose of non-parametric density estimation on linear networks. They further discussed geometric properties of Voronoi tessellations on linear networks. Barr and Schoenberg (2010) focused on the planar case and particular statistical properties.

3 Resample-smoothing of intensity estimators

Barr and Schoenberg (2010) pointed out that when there are abrupt changes in the intensity, kernel-based estimators may yield substantial bias and high variance, and they showed that the Voronoi estimator can alleviate these problems. Unfortunately, Voronoi estimators tend to under-smooth in very dense areas surrounded by nearly empty neighbourhoods. This may be said about adaptive estimators in general; there is a tendency of adapting too much to the particular features of the observed point pattern \mathbf{x} , rather than reflecting the features of the intensity function of the underlying point process X . To see how the under-smoothing, i.e. the over accentuating of local features of the Voronoi intensity estimator occurs, note that for a pattern \mathbf{x} , if $x \in \mathbf{x}$ is located in a very dense part then its Voronoi cell becomes small and, consequently, $\hat{\rho}^V(u) = 1/|\mathcal{V}_x|$ becomes very large for $u \in \mathcal{V}_x$. A further issue with the Voronoi intensity estimator is that its variance tends to be quite large, thus resulting in quite unreliable estimates.

One may further ask whether there are other data-dependent tessellations $\{\mathcal{C}_i\}$, $\bigcup_i \mathcal{C}_i = W$, giving rise to estimators $\hat{\rho}(u) = \sum_i \beta_i \mathbf{1}\{u \in \mathcal{C}_i\}$, $\beta_i > 0$, which perform better than the Voronoi intensity estimator. In addition, an advantage of the kernel estimation approach is arguably in that it generates a smoothly varying intensity estimate, at least when using certain kernels, as opposed to the possibly unnatural “jumps” generated by the Voronoi estimator.

As a remedy for these issues, one suggestion is to follow Barr and Schoenberg (2010) by considering the so-called centroidal Voronoi intensity estimator. A further idea is to introduce a smoothing procedure for $\hat{\rho}^V(\cdot)$, which would reduce the unnaturally extreme peaks while smoothing out the “jumps”. We next propose such a smoothing procedure, which we refer to as *resample-smoothing*.

3.1 Definition of resample-smoothing

Recall the independent thinning operation in Sect. 2.2. We will here focus on the simple case where $p(u) \equiv p \in (0, 1]$, $u \in W$, which is referred to as p -thinning (Chiu et al. 2013, Section 5.1); we identify the case $p = 1$ with the unthinned process X . From Sect. 2.2 we have that

$$\rho(u) = \frac{\rho_{th}(u)}{p}, \quad u \in S,$$

where we recall the intensity $\rho_{th}(\cdot)$ of the thinned process X_p . Hence, dividing by p is exactly what is needed to compensate for the reduced intensity caused by removing points. We exploit this relationship in the following way. Given a point pattern \mathbf{x} and an estimator $\hat{\rho}(\cdot)$ of $\rho(u)$, $u \in W$, fix some $p \in (0, 1]$ and generate $m \geq 1$ independent random patterns, each obtained by randomly thinning the original data pattern \mathbf{x} with retention probability p . This yields thinned patterns $\mathbf{x}_p^1, \dots, \mathbf{x}_p^m$, for each of which the intensity is estimated. We now let the average of these m estimated intensity functions, divided by p , be reported as the final estimate; note the similarities with the approaches considered by Heikkinen and Arjas (1998); Ferreira et al. (2002); Baddeley (2007). The resample-smoothed Voronoi intensity estimator is formally defined as follows.

Definition 2 Consider a point process X in S with intensity function $\rho(\cdot)$. Given some $p \in (0, 1]$ and $m \geq 1$, the *resample-smoothed Voronoi intensity estimator* of $\rho(u)$, $u \in W \subseteq S$, $|W| > 0$, is given by

$$\hat{\rho}_{p,m}^V(u) = \hat{\rho}_{p,m}^V(u; X, W) = \frac{1}{m} \sum_{i=1}^m \frac{\hat{\rho}_i^V(u)}{p}, \quad (4)$$

where

$$\hat{\rho}_i^V(u) = \hat{\rho}^V(u; X_p^i, W) = \sum_{x \in X_p^i} \frac{\mathbf{1}\{u \in \mathcal{V}_x(X_p^i, W)\}}{|\mathcal{V}_x(X_p^i, W)|}$$

is the Voronoi intensity estimator based on the i th thinning X_p^i of $X \cap W$. Note that when $p = 1$, $\hat{\rho}_{p,m}^V(\cdot)$ reduces to $\hat{\rho}^V(\cdot)$ for any $m \geq 1$.

Reflecting on the effect of the thinning procedure, for each thinned version we obtain new Voronoi cells and consequently different locations of the jumps in the corresponding intensity estimate $\hat{\rho}_i^V(\cdot)$. This is what results in the “smoothing” and it is also the remedy for choosing the specific tiling in a possibly wrong/rigid way. Note also that we in fact simply are considering the average of m different estimates of $\rho(\cdot)$.

3.2 Theoretical properties

We next look closer at some statistical properties of resample-smoothed Voronoi intensity estimators. The proofs of all the results presented can be found in the Electronic Supplementary Material (Online Resource 1).

We stress that in the case of the restriction $X \cap W$ of a point process X to a (bounded) region $W \neq S$, the Voronoi cells $\mathcal{V}_x(X, W)$ are different than when $W = S$. Hereby, distributional properties of $\hat{\rho}_{p,m}^V(\cdot)$ may be different depending on how W is chosen.

We start by considering the asymptotic scenario where the number of thinned patterns, $m \geq 1$, in the estimator (4) tends to infinity. Note that by the result below, we have that the limit $\lim_{m \rightarrow \infty} \hat{\rho}_{p,m}^V(u; X, W)$ a.s. exists for a point process X .

Lemma 1 Given fixed $p \in (0, 1]$, for any point pattern $\mathbf{x} \subset W \subseteq S$ we have that $\lim_{m \rightarrow \infty} \hat{\rho}_{p,m}^V(u; \mathbf{x}, W)$ a.s. exists.

3.2.1 Bias

Turning to the first order properties of $\hat{\rho}_{p,m}^V(\cdot)$, we note that

$$\begin{aligned} \int_W \hat{\rho}_{p,m}^V(u) du &= \frac{1}{mp} \sum_{i=1}^m \sum_{x \in X_p^i} \frac{\int_W \mathbf{1}\{u \in \mathcal{V}_x(X_p^i, W)\} du}{|\mathcal{V}_x(X_p^i, W)|} \\ &= \frac{1}{mp} \sum_{i=1}^m N(X_p^i \cap W). \end{aligned} \quad (5)$$

Hence, when $p = 1$ we have preservation of mass, i.e. $\int_W \hat{\rho}_{p,m}^V(u) du = N(X \cap W)$. Taking expectations on both sides in (5), we obtain

$$\mathbb{E} \left[\int_W \hat{\rho}_{p,m}^V(u) du \right] = \frac{1}{m} \sum_{i=1}^m \frac{p \int_W \rho(u) du}{p} = \int_W \rho(u) du, \quad (6)$$

i.e. for any $m \geq 1$ and $p \in (0, 1]$, $\int_W \hat{\rho}_{p,m}^V(u) du$ is an unbiased estimator of $\mathbb{E}[N(X \cap W)]$, and by the law of large numbers, Eq. (5) converges to (6) a.s. as $m \rightarrow \infty$.

Noting that $\mathbb{E}[\hat{\rho}_{p,m}^V(u; X, W)] = \mathbb{E}[\hat{\rho}^V(u; X_p, W)]/p$ for any $p \in (0, 1]$ and $m \geq 1$, we see that $\hat{\rho}_{p,m}^V(u; X, W)$ is unbiased for the estimation of the intensity of X if and only if the original Voronoi intensity estimator is unbiased for the estimation of the intensity of an arbitrary thinning X_p . There is unfortunately not much more to be said without explicitly assuming something about the distributional properties of X .

When X is stationary (see Sect. 2), all Voronoi cells have the same distribution and we may speak of the typical Voronoi cell $\mathcal{V}_o = \mathcal{V}_o(X)$, which satisfies $\mathcal{V}_o \stackrel{d}{=} \theta_{-x}\mathcal{V}_x(X, S)$ for any $x \in X$; here θ_{-x} denotes the transformation/shift such that x is taken to the origin $o \in S$. In particular, we have that $\hat{\rho}_{p,m}^V(u)$ and $\hat{\rho}_{p,m}^V(v)$ have the same distribution for any $u, v \in S$ and it can be shown that unbiasedness holds.

Theorem 1 *For a stationary point process X in $W = S$ with constant intensity $\rho > 0$, the resample-smoothed Voronoi intensity estimator (4) is unbiased for any choice of $p \in (0, 1]$ and $m \geq 1$.*

As our main interest lies in estimating non-constant intensity functions, stationary models are of limited practical interest. We next turn to inhomogeneous Poisson processes in Euclidean spaces.

Theorem 2 *Let X be a Poisson process in $W = S = \mathbb{R}^d$, $d \geq 1$, with intensity function $\rho(u)$, $u \in \mathbb{R}^d$, which satisfies the Lipschitz condition that for some $\mu_u > 0$, $|\rho(v) - \rho(u)| \leq \mu_u \varepsilon$ for $v \in B(u, \varepsilon)$ and $\varepsilon > 0$ sufficiently small; $B(u, \varepsilon)$ denotes the Euclidean ball with centre u and radius $\varepsilon > 0$. Denoting by $C_u(X)$ the Voronoi cell containing $u \in \mathbb{R}^d$, assume further that $m^\kappa := \sup_{u \in \mathbb{R}^d} \mathbb{E}[|C_u(X)|^{-\kappa}] < \infty$ for some $\kappa \geq 1 + 1/d$. Then, for any $u \in \mathbb{R}^d$, $p \in (0, 1]$ and $m \geq 1$,*

$$|\rho(u) - \mathbb{E}[\hat{\rho}_{p,m}^V(u)]| \leq Cp^{-1}(p\rho(u))^{-1/d} \log(p\rho(u))^{2/d}$$

for some $C > 0$ that depends on the intensity. The right hand side tends to 0 as the intensity tends to infinity.

Remark 1 The moment condition and the Lipschitz assumption on ρ can be relaxed to weaker versions and still have the left hand side go to 0, but the rate would be different.

It has been conjectured that the size of the typical cell of a homogeneous Poisson process follows a (generalised) Gamma distribution (see e.g. Chiu et al. 2013); note in particular Lemma 2 below. The moment condition in the statement of the above result, i.e. $m^\kappa < \infty$, would be satisfied if this is indeed the case. Under such a conjectured distribution, Barr and Schoenberg (2010) showed that in the planar case the original Voronoi intensity estimator is ratio-unbiased for a given class of intensity functions.

3.2.2 Variance

Regarding the variance of $\hat{\rho}_{p,m}^V(u)$, the next result shows that by thinning as much as possible we also obtain a variance of the resample-smoothed Voronoi estimator which is close to 0. We see that for cases where the estimator is unbiased we should, in theory, smooth as much as possible, in combination with choosing m as large as possible.

Theorem 3 *Consider a point process X restricted to $W \subseteq S$, where $\hat{\rho}^V(u) = \hat{\rho}^V(u; X, W)$, $u \in W$, has finite variance. Given $p \in (0, 1]$ and $m \geq 1$, the variance of $\hat{\rho}_{p,m}^V(u) = \hat{\rho}_{p,m}^V(u; X, W)$ satisfies*

$$\text{Var}(\hat{\rho}_{p,1}^V(u))/m \leq \text{Var}(\hat{\rho}_{p,m}^V(u)) \leq \text{Var}(\hat{\rho}_{p,1}^V(u))$$

and $\text{Var}(\hat{\rho}_{p,m}^V(u))$ converges as $m \rightarrow \infty$ to the covariance between $\hat{\rho}^V(u; X_p^1, W)/p$ and $\hat{\rho}^V(u; X_p^2, W)/p$, where X_p^1 and X_p^2 are two arbitrary p -thinned versions of X .

Let $m \geq 1$ be fixed. For a bounded $W \subseteq S$ it follows that $\lim_{p \rightarrow 0} \text{Var}(\hat{\rho}_{p,m}^V(u; X, W)) = 0$. Moreover, considering a sequence $W_p \subseteq S$, $p \in (0, 1]$, which increases (in terms of inclusion) as p decreases and satisfies $\mathbb{E}[N(X_p \cap W_p)] = p \int_{W_p} \rho(u) du \rightarrow 0$ as $p \rightarrow 0$, we have that $\lim_{p \rightarrow 0} \text{Var}(\hat{\rho}_{p,m}^V(u; X, W_p)) = 0$.

Turning to the stationary case, from the proof of Theorem 1 (Online Resource 1) we have that the p -thinning X_p of a stationary point process X with intensity $\rho > 0$ is again stationary, but with intensity $p\rho$. For X_p , the distribution $\bar{P}_p(\cdot)$ of the size of the cell that covers u is the same for any $u \in S$ and it is given by [see Last (2010, Section 8) and Schneider and Weil (2008, Theorem 10.4.1.)]

$$\bar{P}_p(A) = p\rho \int_A t P_{|\mathcal{V}_o(X_p)|}(dt), \quad A \subseteq [0, \infty), \quad (7)$$

where $P_{|\mathcal{V}_o(X_p)|}(\cdot)$ is the distribution of the typical cell size. Besides giving us the unbiasedness in Theorem 1, i.e.

$$\mathbb{E}[\hat{\rho}_{p,m}^V(u)] = p^{-1} p\rho \int_0^\infty t^{-1} t P_{|\mathcal{V}_o(X_p)|}(dt) = \rho,$$

the relationship (7) further yields

$$\begin{aligned} \mathbb{E}[\hat{\rho}_{p,1}^V(u)^2] &= \frac{1}{p^2} \int_0^\infty \frac{1}{t^2} \bar{P}_p(dt) = \frac{\rho}{p} \int_0^\infty \frac{1}{t} P_{|\mathcal{V}_o(X_p)|}(dt) \\ &= \frac{\rho}{p} \mathbb{E}[1/|\mathcal{V}_o(X_p)|], \\ \text{Var}(\hat{\rho}_{p,1}^V(u)) &= \frac{\rho}{p} \mathbb{E}[1/|\mathcal{V}_o(X_p)|] - \rho^2. \end{aligned}$$

Through the proof of Theorem 3 (Online Resource 1) we obtain that the variance of $\hat{\rho}_{p,m}^V(u)$ is given by

$$\rho \left(\frac{\mathbb{E}[1/|\mathcal{V}_o(X_p)|]}{p} - \rho \right) \times \frac{1 + (m-1)\text{Corr}(\hat{\rho}^V(u; X_p^1, S), \hat{\rho}^V(u; X_p^2, S))}{m}, \quad (8)$$

where Corr denotes correlation. Unfortunately, we cannot get much further in the general setup; the problem lies in that $P_{|\mathcal{V}_o|}(\cdot)$ typically is not known.

There is, however, one particular case where we can say a bit more and that is for Poisson processes on \mathbb{R} .

Lemma 2 *For a Poisson process on \mathbb{R} with intensity $\rho > 0$, for any $p \in (0, 1]$ and $m \geq 1$ the typical cell size of X_p follows an Erlang/Gamma distribution with shape and rate parameters 2 and $2pp$, respectively. Hence $\text{Var}(\hat{\rho}_{p,m}^V(u)) \leq \text{Var}(\hat{\rho}_{p,1}^V(u)) = \rho^2$.*

Empirically, we have consistently observed that for a large enough m , the variance of $\hat{\rho}_{p,m}^V(u)$ decreases as p decreases, for $u \in W$ located a given distance from the boundary of $W \subseteq S$. As this is partly supported by Theorem 3, we are led to the following conjecture.

Conjecture 1 *For an arbitrary point process X in S and a large enough m , the variance of $\hat{\rho}_{p,m}^V(u)$ is a decreasing function of $p \in (0, 1]$. In particular, if $\hat{\rho}_{p,m}^V(u)$ is unbiased, this means that MISE is decreasing with p .*

3.3 Choosing the smoothing parameters

When using the resample-smoothed Voronoi intensity estimator (4) in practice, one needs to specify the smoothing parameters $m \geq 1$ and $p \in (0, 1]$ prior to finding the intensity estimate. We next discuss how to obtain proper choices for m and p .

3.3.1 Choosing the number of thinnings

Lemma 1 tells us that for a fixed $p \in (0, 1]$ and any point pattern $\mathbf{x} \subset W \subseteq S$, we have that $\hat{\rho}_{p,m}^V(u; \mathbf{x}, W)$ exists a.s. as $m \rightarrow \infty$. The question that remains, however, is for which $m \geq 1$ we are sufficiently close to the limit. In our numerical experiments in Sect. 4 we illustrate that the estimated bias and variance of $\hat{\rho}_{p,m}^V(u)$ do not change significantly for $m \geq 200$. Nevertheless, we propose to fix $m = 400$ and then proceed by finding a proper choice for $p \in (0, 1]$.

3.3.2 Choosing retention probability

The selection of $p \in (0, 1]$ is clearly the more delicate matter here; essentially we are faced with problems similar to those of choosing bandwidths in kernel estimation.

Through our numerical experiments (see Sect. 4) we have found that the choice $p \leq 0.2$ seems to generate the best intensity estimates in the sense that the variance-bias-tradeoff is taken into account by keeping both the bias and variance relatively small. From Sect. 4, Theorem 3 and Conjecture 1 it seems that the smaller the p , the better the estimate. We refer to the choice $m = 400$ and $p \leq 0.2$ as our rule-of-thumb. It should be pointed out that very small values for p may require larger values for m .

We also propose a cross-validation approach to select p when a data-driven approach is preferred to the rule-of-thumb. Recalling a comment in Sect. 2.2 about independent thinnings yielding approximate Poissonian distributional properties of the resulting processes, a natural approach to choosing p when the number of thinned patterns, m , is fixed is to consider Poisson process likelihood cross-validation. This method has a long history in the literature of point processes and has e.g. been frequently used for bandwidth selection in kernel-based estimation (Silverman 1986; Loader 1999). More specifically, given a point pattern $\mathbf{x} = \{x_1, \dots, x_n\} \subset W \subseteq S$ and some fixed $m \geq 1$, we choose the corresponding resampling/retention probability as a maximiser of the cross-validation criterion

$$CV(p) = CV_m(p) = \sum_{i=1}^n \log \hat{\rho}_{p,m}^V(x_i; \mathbf{x} \setminus \{x_i\}, W) - \int_W \hat{\rho}_{p,m}^V(u; \mathbf{x}, W) du, \quad p \in (0, 1]. \quad (9)$$

Note that $\hat{\rho}_{p,m}^V(\cdot; \mathbf{x} \setminus \{x_i\}, W)$ is the *leave-one-out* version of $\hat{\rho}_{p,m}^V(\cdot; \mathbf{x}, W)$, i.e. the resample-smoothed Voronoi intensity estimator based on the reduced sample $\mathbf{x} \setminus \{x_i\}$. Computation of $CV(p)$, $p \in (0, 1]$, can be quite costly. In practice we may ignore the integral term in (9) since it is approximately equal to the number of points in the pattern. Moreover, in practice we calculate $CV(p_j)$, $j = 1, \dots, k$, $0 < p_{j-1} < p_j \leq 1$, sequentially by first generating $X_{p_k}^i$ and then iteratively generating $X_{p_{j-1}}^i = (X_{p_j}^i)_{p_{j-1}/p_j}$, $i = 1, \dots, m$, $j = 2, \dots, k$. Note that for small m the graph of $CV(p)$ may not be smooth and might contain local extrema.

Finally, if the value obtained for p through the cross-validation would deviate too much from the rule-of-thumb, we recommend following the rule-of-thumb; see the log-Gaussian Cox process example in Sect. 4 for a situation where this occurs.

3.4 Large scale data and sparsity

In general, when the number of events, n , of an observed point pattern $\mathbf{x} = \{x_1, \dots, x_n\}$ is very large, it is often natural to consider an adaptive intensity estimator as the scales of intensity likely vary a lot.

It may not be computationally feasible to compute $\hat{\rho}_{p,m}^V(\cdot)$, $p \in (0, 1]$, for an arbitrary $m \geq 1$ (or any other intensity estimator for that matter). An alternative way of exploiting the proposed setup is to consider $\hat{\rho}_{p_0,m}^V(\cdot)$ for some $p_0 \leq 0.2$ and $m = 1$. This means that we would introduce sparsity by only having to generate Voronoi cells for less than 30% of the original number of points. The results in Sect. 4 indicate how good an estimate one would typically obtain. Moreover, if the computation of $\hat{\rho}_{p_0,1}^V(\cdot)$ is reasonably quick, one could generate a further estimate $\hat{\rho}_{p_0,1}^V(\cdot)$ and average over these to obtain $\hat{\rho}_{p_0,2}^V(\cdot)$. One could then continue like this in a stepwise fashion, given a total computation timeframe. This approach could also be useful in machine learning settings (cf. Holmström and Hamalainen 1993); note that $\hat{\rho}_{p,m}^V(\cdot)/n$ is a density estimate for a sample $\mathbf{x} = \{x_1, \dots, x_n\} \subset W$.

4 Numerical experiments

As previously pointed out, we assess our intensity estimation approach numerically, which we choose to do in the Euclidean setting.

In our simulation study, we consider four different types of models with varying degrees of variation in intensity and spatial interaction; clustering, spatial randomness and regularity. For each model we use 500 realisations on $W = [0, 1]^2$ to generate numerical estimates of relevant quantities such as bias, variance, Integrated Variance (IV), Integrated Square Bias (ISB) and Integrated Absolute Bias (IAB) for $\hat{\rho}_{p,m}^V(u)$, $u \in W$; recall that Mean Integrated Square Error (MISE) is obtained as the sum of IV and ISB.

The resample-smoothed Voronoi estimators in the two-dimensional plane and on linear networks were implemented in the R language using the package `spatstat` (Baddeley et al. 2015) and will be released publicly in a future version of `spatstat`. Our simulation experiments and figures were generated using this implementation.

For each model considered, in the Electronic Supplementary Material (Online Resource 2), we provide plots of the estimated bias and variance for $m = 400$ and a range of values of $p \in (0, 1]$, together with the estimated biases and variances obtained through kernel estimation. There, we additionally provide box plots related to point-wise estimation errors.

The overall conclusion is that we clearly reduce the estimation errors by resample-smoothing the Voronoi intensity estimator. Moreover, the cross-validation approach to select-

ing p on average yields slightly poorer intensity estimates than the rule-of-thumb, in particular if the model is clustered. Looking at the box plots in the Electronic Supplementary Material (Online Resource 2), we argue that when e.g. $p = 0.01$ our proposed approach outperforms the two competing kernel estimation approaches, when we are considering clustering or spatial randomness. Under regularity the picture is a bit more varied—the proposed method performs better than the kernel based approaches in terms of extreme over and under estimation. Note that in some situations even a larger p yields similar results.

4.1 Homogeneous Poisson process

Here we consider a homogeneous Poisson process X in $W = [0, 1]^2$ with intensity $\rho = 60$. Table 1 provides estimates of IAB, ISB and IV for $\hat{\rho}_{p,m}^V(u)$, $u \in W$, $m = 200, 300, 400$ and a range of values for p ; recall that we use 500 realisations of X . Indeed, the bias seems fairly stable over the range of values for p and the variance is clearly decreasing with p ; choosing p according to the rule-of-thumb keeps MISE small. For illustrational purposes, in Fig. 1 we provide estimation error plots for one of the realisations, for $p = 0.01$ and $p = 1$ with $m = 400$. One can clearly see the gain of the resample-smoothing. In addition, in the Electronic Supplementary Material (Online Resource 2) we provide plots of the estimated bias and variance for $p = 0.01, 0.1, 0.3, 0.5, 0.7, 1$ and $m = 400$, together with estimation-error-based box plots, and they essentially confirm what has been observed in Table 1.

Turning to the cross-validation approach to selecting p , with $m = 400$, based on 500 realisations of the model we obtain IAB = 3.1, ISB = 13.3 and IV = 177 which are in the range of what one obtains when p is less than 0.3. In Table 2 we further provide the 500 selected values for p and we see that the majority of them fall within the range of our rule-of-thumb.

Comparing with kernel estimation under uniform edge correction, using Poisson likelihood cross-validation (Loader 1999, Sec. 5.3, pp. 87–95) to select the bandwidth, we obtain IAB = 0.24, ISB = 0.11 and IV = 126.05. By instead employing the bandwidth selection method of Cronie and van Lieshout (2018), we obtain IAB = 0.87, ISB = 1.12 and IV = 688.25. Hence, when p is small enough the proposed approach outperforms both kernel approaches in terms of MISE.

4.2 Inhomogeneous Poisson process

More interestingly, we next consider 500 realisations of an inhomogeneous Poisson process X in $W = [0, 1]^2$ with intensity $\rho(x, y) = |10 + 90 \sin(16x)|$; the expected total point count is 58.6. Table 3 provides estimates of IAB, ISB

Table 1 Estimates of IAB, ISB and IV for $\hat{\rho}_{p,m}^V(u)$, $u \in W = [0, 1]^2$, $m = 200, 300, 400$ and a sequence of p , based on 500 realisations of a homogeneous Poisson process in $W = [0, 1]^2$ with intensity $\rho = 60$

	IAB			ISB			IV		
m									
p	200	300	400	200	300	400	200	300	400
.01	2.21	2.20	2.21	6.86	6.85	6.91	97.83	85.04	77.99
.03	4.71	4.69	4.68	30.63	30.42	30.35	92.59	87.18	85.34
.05	5.63	5.64	5.64	43.09	43.1	43.17	108.05	105.11	100.95
.1	5.7	5.7	5.7	43.5	43.2	43.0	158.4	154.8	152.5
.2	4.6	4.6	4.6	28.4	28.5	28.4	264.1	260.3	257.9
.3	3.9	3.9	3.9	22.5	22.2	22.2	375.3	370.6	368.8
.4	3.5	3.5	3.5	19.7	19.6	19.6	490.6	488.8	487.8
.5	3.2	3.2	3.2	18.1	18.1	18.1	672.0	623.9	622.9
.6	3.0	3.0	3.0	17.1	17.1	17.0	781.9	779.4	779.0
.7	2.9	2.9	2.9	16.5	16.5	16.5	960.0	958.7	958.8
.8	2.9	2.9	2.9	16.0	16.0	16.0	1172.2	1171.8	1171.1
.9	2.9	2.9	2.9	15.8	15.8	15.8	1422.2	1419.6	1418.9
1	2.9	2.9	2.9	15.8	15.8	15.8	1733.2	1733.2	1733.2

Fig. 1 Estimation error plots for a realisation of a homogeneous Poisson process X in $W = [0, 1]^2$ with intensity $\rho = 60$. Left: $p = 0.01$ and $m = 400$. Right: $p = 1$. The underlying point pattern has been superimposed in all plots

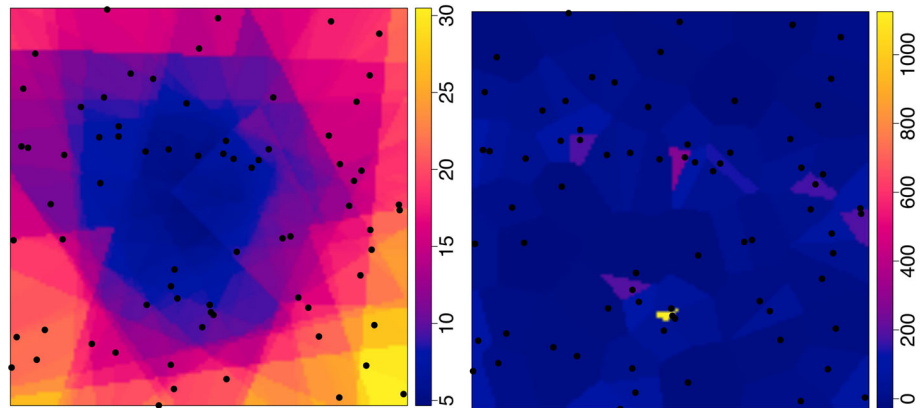


Table 2 Cross-validation selections of p for $m = 400$ in a geometric sequence, based on 500 realisations of a homogeneous Poisson process in $W = [0, 1]^2$ with intensity $\rho = 60$

p	0.01	0.02	0.03	0.07	0.12	0.23	0.43	0.80
Frequency	214	113	60	30	23	30	24	6

and IV for $\hat{\rho}_{p,m}^V(u)$, $u \in W$, $m = 200, 300, 400$ and a range of values for p . Moreover, in Fig. 2 we provide estimation error plots for one of the realisations, for $p = 0.01$ and $p = 1$ with $m = 400$, and in the Electronic Supplementary Material (Online Resource 2), we provide plots of the estimated bias and variance for $p = 0.01, 0.1, 0.3, 0.5, 0.7, 1$ and $m = 400$, together with estimation-error-based box plots, and they likewise indicate the advantage of resample-smoothing.

Turning to the cross-validation approach to selecting p , based on $m = 400$ and 500 realisations of the model, we obtain $\text{IAB} = 25.3$, $\text{ISB} = 867.4$ and $\text{IV} = 174.8$, with

the majority of the selected p 's coinciding with the rule-of-thumb (see Table 4).

Hence, the conclusions here are essentially the same as for the homogeneous Poisson process in Sect. 4.1, with the main difference arguably being that inhomogeneity enforces slightly harder thinning in the cross-validation.

Comparing with kernel estimation under uniform edge correction, using Poisson likelihood cross-validation (Loader 1999, Sec. 5.3, pp. 87–95) to select the bandwidth, we obtain $\text{IAB} = 25.16$, $\text{ISB} = 853.24$ and $\text{IV} = 158.00$. By instead employing the bandwidth selection method of Cronie and van Lieshout (2018), we obtain $\text{IAB} = 24.43$, $\text{ISB} = 797.02$ and $\text{IV} = 636.63$. Thus, for $p < 0.1$, the proposed approach outperforms both kernel methods in terms of MISE. In particular, for both the homogeneous and the inhomogeneous Poisson process examples, when $p \leq 0.4$ our proposed method shows a better performance in terms of MISE than the kernel approach with the bandwidth selection method of Cronie and van Lieshout (2018).

Table 3 Estimates of IAB, ISB and IV for $\hat{\rho}_{p,m}^V(u)$, $u \in W = [0, 1]^2$, $m = 200, 300, 400$ and a sequence of p , based on 500 realisations of an inhomogeneous Poisson process on $W = [0, 1]^2$ with intensity $\rho(x, y) = |10 + 90 \sin(16x)|$

IAB			ISB			IV			
m									
p	200	300	400	200	300	400	200	300	400
.01	25.3	25.3	25.3	867.0	866.8	867.1	96.7	84.7	79.8
.03	25.4	25.5	25.4	881.7	881.8	882.2	93.6	85.1	80.9
.05	25.5	25.5	25.5	891.7	891.8	891.9	104.8	99.4	97.6
.1	25.6	25.6	25.6	892.3	891.8	891.7	154.2	150.1	147.6
.2	25.5	25.5	25.5	882.8	883.2	883.3	249.1	247.3	245.6
.3	25.6	25.5	25.5	881.5	881.5	881.5	360.1	356.3	356.2
.4	25.5	25.5	25.5	878.8	879.0	879.0	479.9	477.2	475.0
.5	25.5	25.5	25.5	872.6	872.5	872.6	609.8	609.6	609.8
.6	25.4	25.4	25.4	862.7	862.7	862.7	762.6	764.3	764.1
.7	25.2	25.2	25.2	849.9	850.0	850.0	952.0	948.3	949.0
.8	25.0	25.0	25.0	835.1	834.8	834.8	1171.9	1172.3	1172.1
.9	24.7	24.7	24.7	817.7	817.6	817.6	1440.1	1440.9	1440.0
1	24.4	24.4	24.4	799.3	799.3	799.3	1783.8	1783.8	1783.8

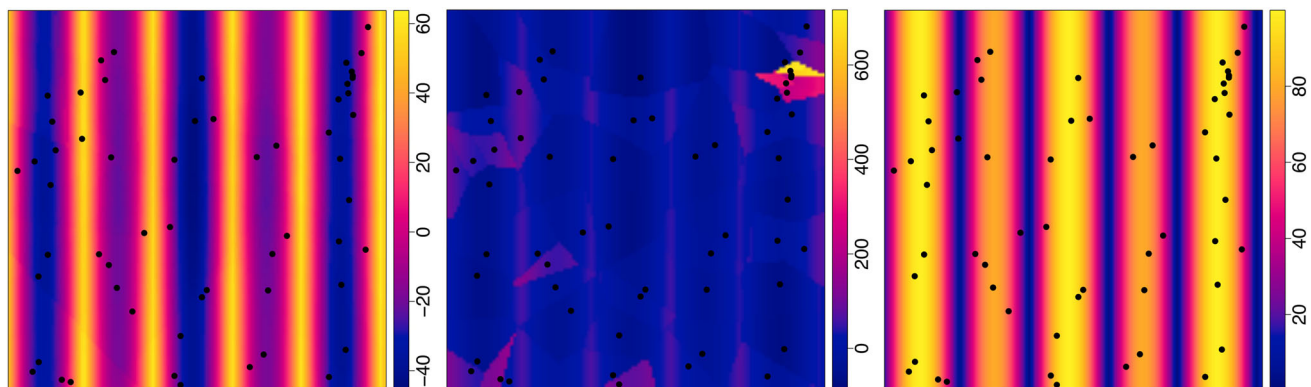


Fig. 2 True intensity and estimation error plots for a realisation of an inhomogeneous Poisson process on $W = [0, 1]^2$ with intensity $\rho(x, y) = |10 + 90 \sin(16x)|$. Left: $p = 0.01$ and $m = 400$. Middle: $p = 1$. Right: true intensity. The underlying point pattern has been superimposed in all plots

Table 4 Cross-validation selections of p in a geometric sequence for $m = 400$, based on 500 realisations of an inhomogeneous Poisson process in $W = [0, 1]^2$ with intensity $\rho(x, y) = |10 + 90 \sin(16x)|$

p	0.01	0.02	0.03	0.07	0.12	0.23	0.43	0.80
Frequency	221	116	66	34	12	25	17	9

4.3 Log-Gaussian Cox process

Turning to the scenario where the underlying point process exhibits clustering, we next consider 500 realisations of a log-Gaussian Cox process X in $W = [0, 1]^2$ where the driving Gaussian random field has the mean function $(x, y) \mapsto \log(40|\sin(20x)|)$ and covariance function $((x_1, y_1), (x_2, y_2)) \mapsto 2 \exp\{-\|(x_1, y_1) - (x_2, y_2)\|/0.1\}$. Hereby, the intensity is given by $\rho(x, y) = 40|\sin(20x)|e^1$. Table 5 provides estimates of IAB, ISB and IV for $\hat{\rho}_{p,m}^V(u)$,

$u \in W$, $m = 200, 300, 400$ and a range of values of p . We see that the rule-of-thumb, i.e. $p \leq 0.2$, seems to be the preferable choice. In Fig. 3 we provide estimation error plots for one of the realisations, for $p = 0.01$ and $p = 1$ with $m = 400$, and in the Electronic Supplementary Material (Online Resource 2), we provide plots of the estimated bias and variance for $p = 0.01, 0.1, 0.3, 0.5, 0.7, 1$ and $m = 400$ together with estimation-error-based box plots. Here it becomes visually clear that the resample-smoothing is improving the estimation quite significantly.

The cross-validation approach to selecting p , based on $m = 400$ and 500 realisations of the model, yields IAB = 26.3, ISB = 948.2 and IV = 23,580.7, which may be comparable to the choice $p \approx 0.7$. In Table 6 we further provide the 500 selected values for p . The phenomenon that too little smoothing tends to be applied (p is mainly chosen large) is not extremely surprising; as our cross-validation

Table 5 Estimates of IAB, ISB and IV for $\hat{\rho}_{p,m}^V(u)$, $u \in W = [0, 1]^2$, $m = 200, 300, 400$ and a sequence of p , based on 500 realisations of a log-Gaussian Cox process in $W = [0, 1]^2$ with mean function $(x, y) \mapsto \log(40|\sin(20x)|)$ and covariance function $((x_1, y_1), (x_2, y_2)) \mapsto 2 \exp\{-\|(x_1, y_1) - (x_2, y_2)\|/0.1\}$ for the driving random field

	IAB			ISB			IV ($\times 10^2$)		
m									
p	200	300	400	200	300	400	200	300	400
.01	29.2	29.2	29.2	1144.2	1144.6	1144.4	10.0	10.2	10.1
.03	29.6	29.7	29.7	1186.7	1186.6	1187.1	18.0	17.5	17.3
.05	29.8	29.7	29.7	1199.2	1199.5	1199.1	26.5	26.5	26.7
.1	29.5	29.5	29.5	1181.5	1181.9	1180.9	48.8	48.8	48.7
.2	28.8	28.8	28.8	1127.3	1127.4	1127.3	87.8	87.2	88.0
.3	28.2	28.2	28.2	1081.4	1081.7	1081.6	123.8	122.6	123.1
.4	27.6	27.6	27.6	1038.8	1039.2	1039.4	153.2	153.0	152.6
.5	27.1	27.1	27.1	1000.1	999.6	999.7	181.3	182.2	182.0
.6	26.5	26.5	26.5	963.9	963.7	963.5	212.4	212.5	212.1
.7	26.0	26.0	26.0	930.5	930.4	930.6	243.1	243.0	243.2
.8	25.6	25.6	25.6	901.1	900.6	900.7	278.8	279.2	279.3
.9	25.2	25.2	25.2	874.4	874.3	874.2	321.4	321.5	320.9
1	24.7	24.7	24.7	852.3	852.3	852.3	371.4	371.4	371.4

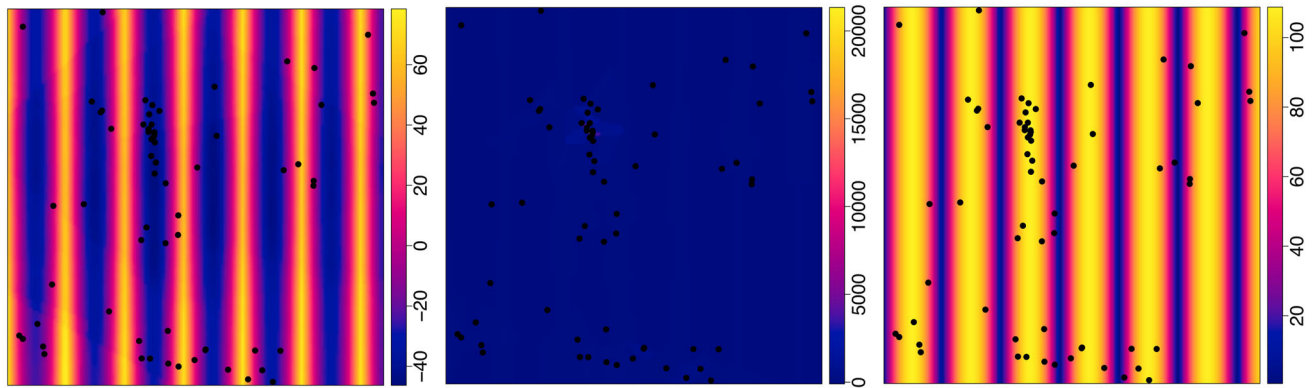


Fig. 3 True intensity and estimation error plots for a realisation of a log-Gaussian Cox process in $W = [0, 1]^2$ with mean function $(x, y) \mapsto \log(40|\sin(20x)|)$ and covariance function $((x_1, y_1), (x_2, y_2)) \mapsto$

$2 \exp\{-\|(x_1, y_1) - (x_2, y_2)\|/0.1\}$ for the driving random field. Left: $p = 0.01$ and $m = 400$. Middle: $p = 1$. Right: true intensity. The underlying point pattern has been superimposed in all plots

Table 6 Cross-validation selections of p in a geometric sequence for $m = 400$, based on 500 realisations of a log-Gaussian Cox process in $W = [0, 1]^2$ with mean function $(x, y) \mapsto \log(40|\sin(20x)|)$ and covariance function $((x_1, y_1), (x_2, y_2)) \mapsto 2 \exp\{-\|(x_1, y_1) - (x_2, y_2)\|/0.1\}$ for the driving random field

p	0.01	0.02	0.03	0.07	0.12	0.23	0.43	0.80
Frequency	7	4	6	1	10	22	207	243

approach is based on a Poisson process likelihood function, it treats a realisation \mathbf{x} of X as a realisation of a Poisson process which has the corresponding realisation of the driving (random) intensity field as intensity function. In other words, it tries to perform state estimation, i.e. it tries to reconstruct each realisation of the driving intensity field through \mathbf{x} . This phenomenon, and that the Poisson process likelihood cross-validation approach is not performing well for clustered inhomogeneous point processes, has previously been

observed in the context of kernel intensity estimation (Cronie and van Lieshout 2018). Hence, if one suspects that there is clustering in addition to inhomogeneity, or if the cross-validation generates large values for p , then it is wiser to stick with the proposed rule-of-thumb, $p \leq 0.2$. In fact, cross-validation-generated deviations from the rule-of-thumb may be seen as a possible indication of clustering or inhibition.

Comparing with kernel estimation under uniform edge correction, using Poisson likelihood cross-validation (Loader 1999, Sec. 5.3, pp. 87–95) to select the bandwidth, we obtain $\text{IAB} = 27.75$, $\text{ISB} = 1031.03$ and $\text{IV} = 9952.85$. By instead employing the bandwidth selection method of Cronie and van Lieshout (2018), we obtain $\text{IAB} = 28.97$, $\text{ISB} = 1117.94$ and $\text{IV} = 3856.79$. We see that our proposed method outperforms both of the kernel-based approaches in terms of MISE when p is small enough. Note, in particular, that in terms of MISE it outperforms the kernel approach with the bandwidth

selection based on the likelihood cross-validation approach when $p \leq 0.2$.

4.4 Thinned simple sequential inhibition point process

To study inhomogeneity in combination with inhibition, we consider a simple sequential inhibition point process in $W = [0, 1]^2$ with a total point count of 450 and inhibition distance 0.3, which we thin according to the retention probability function $p(x, y) = \mathbf{1}\{x < 1/3\}|x - 0.02| + \mathbf{1}\{1/3 \leq x < 2/3\}|x - 0.5| + \mathbf{1}\{x \geq 2/3\}|x - 0.95|$, $x, y \in W$. This results in an inhomogeneous point process with intensity $\rho(x, y) = 450p(x, y)$, which yields an expected total point count of 53.6. Table 7 provides estimates of IAB, ISB and IV for $\hat{\rho}_{p,m}^V(u)$, $u \in W$, $m = 200, 300, 400$ and a range

of values for p . Just as for the previous models, we argue that p should be chosen within the range of the rule-of-thumb.

In Fig. 4 we provide estimation error plots for one of the realisations, for $p = 0.01$ and $p = 1$ with $m = 400$. Plots of the estimated bias and variance, for $p = 0.01, 0.1, 0.3, 0.5, 0.7, 1$ and $m = 400$ together with estimation-error-based box plots can be found in the Electronic Supplementary Material (Online Resource 2). Also here the improvements caused by the resample-smoothing are visually clear.

The cross-validation approach to selecting p based on $m = 400$ and 500 realisations of the model yields IAB = 26.5, ISB = 1033.6 and IV = 508.1, which is comparable to choosing $p \approx 0.5$. Moreover, Table 8 lists the selected values for p and we see that they tend to be either very large or very small. It thus seems that approximately half of the time the

Table 7 Estimates of IAB, ISB and IV for $\hat{\rho}_{p,m}^V(u)$, $u \in W = [0, 1]^2$, $m = 200, 300, 400$ and a sequence of p , based on 500 realisations of an independently thinned simple sequential inhibition process in $W = [0, 1]^2$ with intensity $\rho(x, y) = 450p(x, y)$, $p(x, y) = \mathbf{1}\{x < 1/3\}|x - 0.02| + \mathbf{1}\{1/3 \leq x < 2/3\}|x - 0.5| + \mathbf{1}\{x \geq 2/3\}|x - 0.95|$, $x, y \in W$

p	IAB			ISB			IV		
	m			m			m		
	200	300	400	200	300	400	200	300	400
.01	31.9	31.9	31.9	1458.2	1458.2	1458.4	81.5	69.6	62.4
.03	32.3	32.3	32.3	1493.9	1493.7	1493.4	69.2	65.4	63.0
.05	32.4	32.4	32.4	1511.2	1510.9	1510.4	78.4	75.1	72.2
.1	32.4	32.4	32.4	1502.2	1502.7	1502.2	109.4	105.9	103.4
.2	31.2	31.2	31.2	1385.7	1385.2	1384.5	176.2	173.8	172.2
.3	29.2	29.2	29.2	1223.6	1223.0	1222.8	253.4	251.2	250.3
.4	27.0	27.0	27.0	1060.4	1060.7	1060.3	348.8	345.3	345.3
.5	25.0	25.0	25.0	919.5	919.8	920.6	457.3	455.6	454.1
.6	23.1	23.1	23.1	803.3	803.3	803.0	584.4	582.7	581.9
.7	21.5	21.5	21.5	707.9	707.7	707.8	734.2	733.9	732.8
.8	20.0	20.1	20.1	628.5	628.9	629.1	916.3	914.2	913.4
.9	18.9	18.9	18.9	567.2	567.5	567.7	1120.5	1118.5	1117.5
1	24.7	24.7	24.7	852.3	852.3	852.3	1382.4	1382.4	1382.4

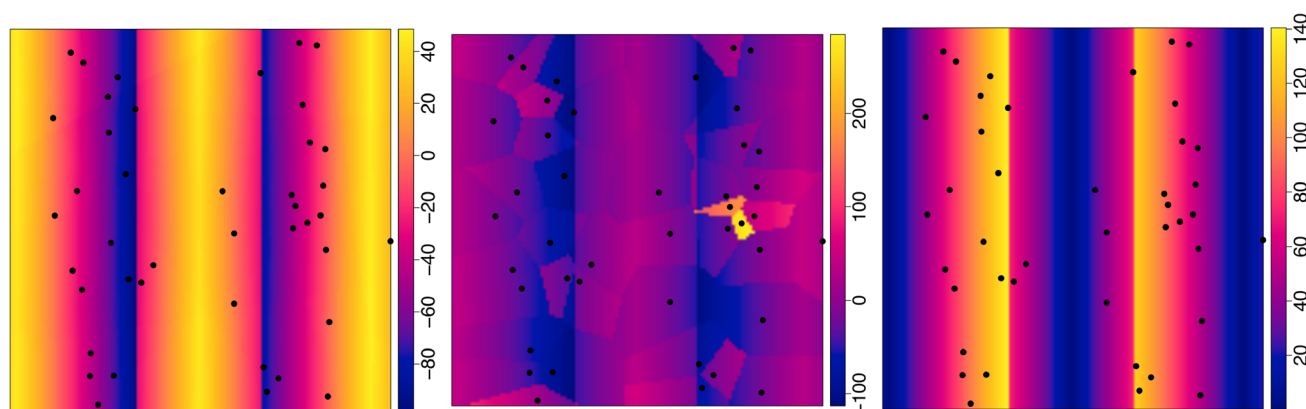


Fig. 4 True intensity and estimation error plots for a realisation of an independently thinned simple sequential inhibition process in $W = [0, 1]^2$ with intensity $\rho(x, y) = 450p(x, y)$, $p(x, y) = \mathbf{1}\{x < 1/3\}|x - 0.02| + \mathbf{1}\{1/3 \leq x < 2/3\}|x - 0.5| + \mathbf{1}\{x \geq 2/3\}|x - 0.95|$,

$x, y \in W$. Left: $p = 0.01$ and $m = 400$. Middle: $p = 1$. Right: true intensity. The underlying point pattern has been superimposed in all plots

Table 8 Table Cross-validation selections of p in a geometric sequence for $m = 400$, based on 500 realisations of an independently thinned simple sequential inhibition process in $W = [0, 1]^2$ with intensity $\rho(x, y) = 450p(x, y)$, $p(x, y) = \mathbf{1}\{x < 1/3\}|x - 0.02| + \mathbf{1}\{1/3 \leq x < 2/3\}|x - 0.5| + \mathbf{1}\{x \geq 2/3\}|x - 0.95|$, $x, y \in W$

p	0.01	0.02	0.03	0.07	0.12	0.23	0.43	0.80
Frequency	142	83	22	8	3	8	79	155

cross-validation performs as it should do and approximately half of the time it chooses p too large.

Comparing with kernel estimation under uniform edge correction, using Poisson likelihood cross-validation (Loader 1999, Sec. 5.3, pp. 87–95) to select the bandwidth, we obtain $IAB = 20.5$, $ISB = 663.94$ and $IV = 485.48$. By instead employing the bandwidth selection method of Cronie and van Lieshout (2018), we obtain $IAB = 23.97$, $ISB = 860.67$ and $IV = 308.47$. We see that the proposed approach performs slightly poorer than kernel approaches.

5 Data analysis

We next apply our proposed intensity estimator (4) to two real datasets, in two types of spaces. We first study a linear network dataset of traffic accidents in an area of Houston, USA, and then a planar dataset of spatial locations of Finnish pines.

5.1 Houston motor vehicle traffic accidents

The dataset consists of motor vehicle traffic accident locations in a given area of Houston, USA, during the month of April 1999. The linear network L describing the road network in question (see Fig. 5) has a total length of 708,301.7 feet, and has 187 vertices (road intersections) with a maximum vertex degree of 4, and 253 line segments, i.e. pieces of streets connecting the intersections.

Fig. 5 Left: motor vehicle traffic accidents in an area of Houston, US, during April, 1999. Right: resample-smoothed Voronoi intensity estimate for $m = 400$ and $p = 0.20$

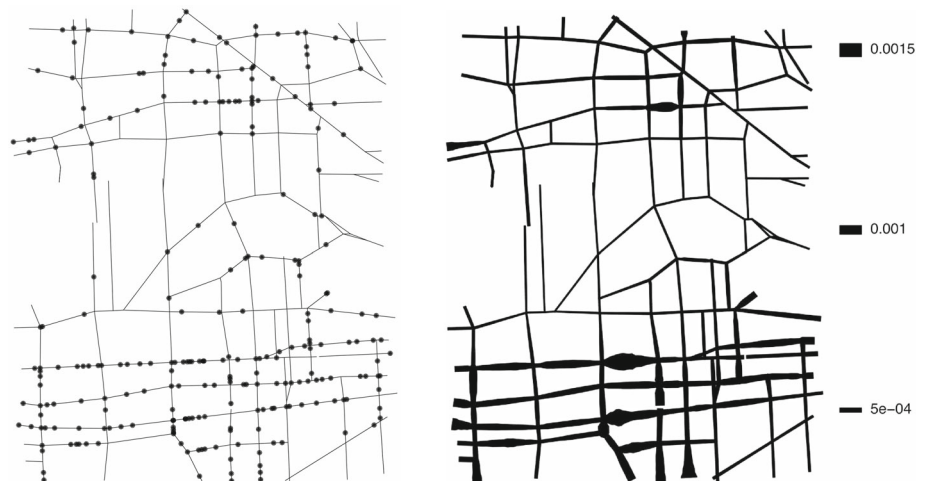


Table 9 Houston motor vehicle traffic accidents: Cross-validation selected values for p , based on the sequence $m = 100, 150, \dots, 400$

m	100	150	200	250	300	350	400
p	0.15	0.15	0.15	0.20	0.15	0.20	0.20

Figure 5 (left) shows the reference points of the 249 accidents over the street network. The data have been collected by individual police departments in the Houston metropolitan area and compiled by the Texas Department of Public Safety. The compiled data have been obtained by the Houston-Galveston Area Council and then geocoded by N. Levine. Between 1999 and 2001, in the eight-county region considered, there were 252,241 serious accidents, with an average of 84,080 per year. From these accidents, 1882 were person related. See Levine (2006, 2009) for details.

In Fig. 5 (right) we also provide the resample-smoothed Voronoi intensity estimate obtained for $m = 400$ and $p = 0.20$. The specific choice $p = 0.20$ has been motivated by the rule-of-thumb $p \leq 0.2$ and Table 9, which shows the selected values for $p \in (0, 1]$ obtained by carrying out cross-validation for the sequence $m = 100, 150, \dots, 400$. We see that selected values for p are given by either 0.15 or 0.20.

Visually, there seems to be a good correspondence between the observed pattern and the obtained estimate. Note that for bigger values of p , in the right panel of Fig. 5 we would have obtained more significant blobs in the parts corresponding to the dense parts in the left panel of Fig. 5.

5.2 Finnish pines

The dataset, which consists of the locations of 126 pine saplings in a Finnish forest, within a rectangular window $W = [-5, 5] \times [-8, 2]$ (metres), can be found in the R package *spatstat* (Baddeley et al. 2015). It was recorded by S. Kellomaki, Faculty of Forestry, University of Joensuu, Fin-

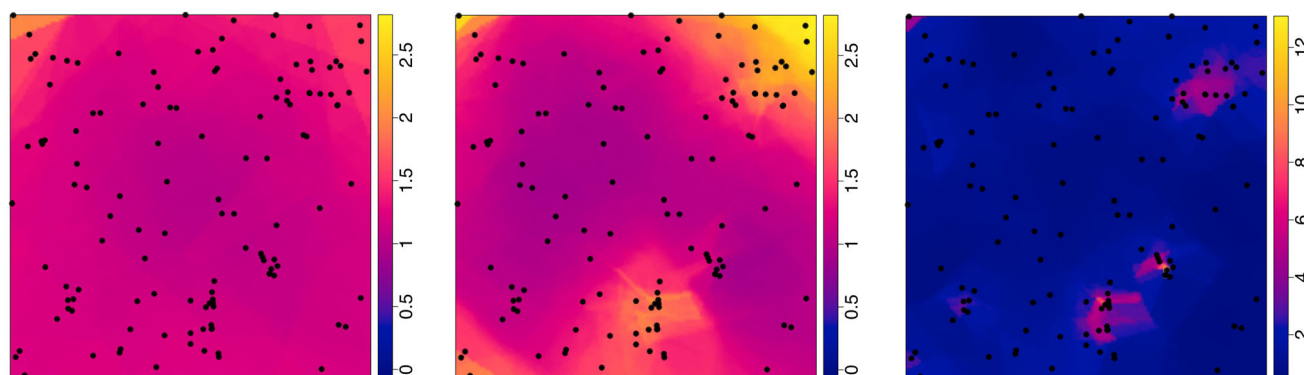


Fig. 6 The estimate $\hat{\rho}_{p,m}^V(u)$, $u \in W$, $m = 400$, for $p = 0.01$ (left), $p = 0.1$ (middle) and $p = 0.45$ (right), together with the locations of 126 pine saplings in a Finnish forest, within a rectangular window $W = [-5, 5] \times [-8, 2]$ (metres)

Table 10 Finish pines: cross-validation selected values for p , based on the sequence $m = 100, 150, \dots, 400$

m	100	150	200	250	300	350	400
p	0.65	0.50	0.50	0.50	0.45	0.50	0.45

land, and further processed by A. Penttinen, Department of Statistics, University of Jyväskylä, Finland.

In Fig. 6 we illustrate the estimate $\hat{\rho}_{p,m}^V(u)$, $u \in W$, $m = 400$, for $p = 0.01$, $p = 0.1$ and $p = 0.45$, together with the locations of the saplings. We further provide the cross-validation results for the sequence $m = 100, 150, \dots, 400$ in Table 10; it suggests the choice $p = 0.45$. We argue that $p = 0.01$ and $p = 0.1$ result in pretty similar intensity maps and they better respect the global features of the data than $p = 0.45$.

6 Discussion and future work

We have proposed a general approach for resampling, or additional smoothing, of Voronoi intensity estimators. It is based on averaging over intensity estimators generated by a set of thinned samples. We believe that its strength lies in that it filters out sporadic/local features in order to accentuate the structural information contained in the sample. In addition, viewing the reciprocal of a point's Voronoi cell size as a type of kernel (cf. van Lieshout 2012), centred at the point, each time we thin the pattern we change the support of that kernel. Having averaged over the thinned estimators, in essence we end up using an “average” support for each such kernel.

In order to determine how much smoothing, i.e. thinning, should be applied, we have proposed both a rule-of-thumb ($m = 400$ and $p \leq 0.2$) and a data-driven cross-validation approach. We have observed that for Poisson and log Gaussian Cox processes, by using resample-smoothed Voronoi intensity estimation together with our rule-of-thumb, we

outperform kernel estimation in terms of Mean Integrated Square Error (MISE) and point-wise over-/under-estimation, based on the state-of-the-art in bandwidth selection. The over-/under-estimation has been illustrated by means of point-wise estimation error box plots which can be found in the Electronic Supplementary Material (Online Resource 2). For regular point process models the picture seems to be a bit more varied—our proposed approach outperforms the kernel approaches in terms of over-/under-estimation and performs slightly poorer in terms of MISE. In essence one could say that if we employ the expected supremum distance to compare the functions then the new method outperforms the kernel method. For the expected L_2 distance, reflected by MISE, however, this is not true for the regular setting.

The performance of the proposed estimator depends on the tuning parameters p and m . The guidelines for choosing p and m have been based on the present examples with a sample size of roughly $n = 60$. In particular, a combination of a smaller sample size and a very small choice of p may call for an increase of m . This should be computationally feasible since each thinned pattern then will consist of very few points and the corresponding Voronoi tessellation will be fast to compute.

It should be noted that we alternatively may employ some retention probability function $p(u)$, $u \in W$, other than $p(u) \equiv p \in (0, 1]$. It is, however, not clear what the benefits of such a change would be, other than possibly decreasing the computational time. Also, how to make a good choice for the function $p(\cdot)$ is not evident.

6.1 Future work and extensions

It would be relevant and interesting to study the proposed setup when we replace the Voronoi tessellation by some other tessellation, generated by the point pattern in question. One such example is provided by Delaunay tessellations, as they enjoy more tractable distributional properties in Euclidean

spaces. Another idea is to consider some other adaptive intensity estimator, e.g. nearest neighbour estimators (Silverman 1986; van Lieshout 2012). Another relevant idea might be applying the resample-smoothing procedure to adaptive kernel estimators (Davies and Hazelton 2010; Davies et al. 2018).

Further possible extensions are discussed below.

6.1.1 Sequential resample-smoothing

Since choosing the smoothing parameter $p \in (0, 1]$ according to the cross-validation approach in Sect. 3.3 can be quite computationally demanding, and thereby also time consuming, we propose an alternative and simpler version of the estimator in (4).

Definition 3 Given some $\mathbf{p}_m = (p_1, \dots, p_m) \in (0, 1]^m$, $m \geq 1$, the *sequentially resample-smoothed Voronoi intensity estimator* of the intensity $\rho(u)$, $u \in W \subseteq S$, $|W| > 0$, of the underlying point process X is defined as

$$\hat{\rho}_{\mathbf{p}_m}^V(u) = \hat{\rho}_{\mathbf{p}_m}^V(u; X, W) = \sum_{j=1}^m \frac{\hat{\rho}^V(u; X_{p_j}, W)}{mp_j}, \quad u \in W,$$

where X_{p_1}, \dots, X_{p_m} is a sequence of independent thinnings of X , with the respective retention probabilities p_j , $j = 1, \dots, m$. In particular, $\hat{\rho}_{p,m}^V(\cdot) = \hat{\rho}_{(p,\dots,p)}^V(\cdot)$.

The challenge here is clearly how to choose the sequence \mathbf{p}_m ; we have seen that more weight clearly should be put on smaller retention probability values so an equally spaced grid over $(0, 1]$ may not be the best choice. By proposing some stepwise sequencing of $(0, 1]$, where we at each step $m \geq 1$ obtain some $\mathbf{p}_m = (p_1, \dots, p_m) \in (0, 1]^m$, one could keep going until $\sup_{u \in W} |\hat{\rho}_{\mathbf{p}_m}^V(u) - \hat{\rho}_{\mathbf{p}_{m+1}}^V(u)| < \epsilon$ or $\sup_{u \in W} |\hat{\rho}_{\mathbf{p}_m}^V(u) - \hat{\rho}_{\mathbf{p}_{m+1}}^V(u)| / \hat{\rho}_{\mathbf{p}_m}^V(u) < \epsilon$ for some predefined $\epsilon > 0$.

6.1.2 Edge correction in the linear network case

Although we have neglected edge effects here, it still seems that the smoothing takes care of a significant part of the edge effects (Chiu et al. 2013). But, as noted in the data analysis, even after applying the smoothing there may be a need for edge correction (Cronie and Särkkä 2011; Baddeley et al. 2015). In the case where X is sampled on L , and is a subset of a process on a larger network, in which L is a sub-network, edge effects come into play since the points closest to the boundary have their Voronoi cells cut off through the mapping/sampling of L and the points. In Definition 4 we propose an edge correction approach, which could be viewed as a version of Ripley's edge correction idea.

Definition 4 Given a point pattern \mathbf{x} on a linear network L , for each boundary point $u \in \partial L$ of $L \subset S$, first find its closest neighbour $x_u = \arg \min_{x \in \mathbf{x}} d(u, x)$ in terms of the shortest path distance $d(\cdot, \cdot)$. If $\beta_u = \min_{x \in \mathbf{x} \setminus \{x_u\}} d(x_u, x)/2 - d(u, x_u) > 0$, extend L by a new (set of) non-overlapping edge(s) connected to the node u , with total length β_u . Denote the resulting extended network by $\tilde{L}(\mathbf{x})$ and treat \mathbf{x} as a linear network point pattern on/restricted to $\tilde{L}(\mathbf{x})$. The edge corrected resample-smoothed Voronoi intensity estimate is given by $\hat{\rho}_{p,m}^V(u; \mathbf{x}, L) = \hat{\rho}_{p,m}^V(u; \mathbf{x}, \tilde{L}(\mathbf{x}))$ for $u \in W$. Note that $p = 1$ results in an edge corrected version of $\hat{\rho}^V(\cdot)$.

Acknowledgements The authors are grateful to the editor and two referees for useful comments. M.M. Moradi gratefully acknowledges funding from the European union through the GEO-C project (H2020-MSCA-ITN-2014, Grant Agreement Number 642332, <http://www.geo-c.eu/>); J. Mateu is partially funded by Grants MTM2016-78917-R from the Spanish Ministry of Science and Education, and P1-1B2015-40 from University of Jaume I. Ege Rubak was supported by The Danish Council for Independent Research | Natural Sciences, Grant DFF—7014–00074 “Statistics for point processes in space and beyond”; by a six-month visiting position at Curtin University; and by the Centre for Stochastic Geometry and Advanced Bioimaging, funded by Grant 8721 from the Villum Foundation. Adrian Baddeley was funded by the Australian Research Council, Discovery Grants DP1301002322 and DP130104470. We thank Ned Levine for kindly providing us with the dataset on Houston vehicle traffic accidents as well as helpful discussions on such data. We also thank David Cohen for fruitful discussions.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Ang, Q.W., Baddeley, A., Nair, G.: Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. *Scand. J. Stat.* **39**(4), 591–617 (2012)
- Baddeley, A.: Validation of statistical models for spatial point patterns. In: Babu, J., Feigelson, E. (eds) *Statistical Challenges in Modern Astronomy IV*, Astronomical Society of the Pacific, San Francisco, California, USA, Astronomical Society of the Pacific, Conference Series, vol. 371, pp. 22–38 (2007)
- Baddeley, A., Rubak, E., Turner, R.: *Spatial Point Patterns: Methodology and Applications* with R. CRC Press, Boca Raton (2015)
- Barr, C.D., Schoenberg, F.P.: On the Voronoi estimator for the intensity of an inhomogeneous planar Poisson process. *Biometrika* **97**(4), 977–984 (2010)
- Berman, M., Diggle, P.: Estimating weighted integrals of the second-order intensity of a spatial point process. *J. R. Stat. Soc. Ser. B* **51**, 81–92 (1989)
- Borruso, G.: Network density estimation: analysis of point patterns over a network. In: *Computational Science and Its Applications—ICCSA 2005*, Springer, pp. 126–132 (2005)
- Borruso, G.: Network density and the delimitation of urban areas. *Trans. GIS* **7**, 177–191 (2003)

- Borruso, G.: Network density estimation: a GIS approach for analysing point patterns in a network space. *Trans. GIS* **12**(3), 377–402 (2008)
- Brown, G.S.: Point density in stems per acre. Forest Research Institute, New Zealand Forest Service (1965)
- Chiu, S.N., Stoyan, D., Kendall, W.S., Mecke, J.: *Stochastic Geometry and Its Applications*. Wiley, Hoboken (2013)
- Cronie, O., Särkkä, A.: Some edge correction methods for marked spatio-temporal point process models. *Comput. Stat. Data Anal.* **55**(7), 2209–2220 (2011)
- Cronie, O., van Lieshout, M.N.M.: A non-model-based approach to bandwidth selection for kernel estimators of spatial intensity functions. *Biometrika* **105**(2), 455–462 (2018)
- Daley, D.J., Vere-Jones, D.: *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*, 2nd edn. Springer, New York (2008)
- Davies, T.M., Baddeley, A.: Fast computation of spatially adaptive kernel estimates. *Stat. Comput.* **28**(4), 937–956 (2018)
- Davies, T.M., Hazelton, M.L.: Adaptive kernel estimation of spatial relative risk. *Stat. Med.* **29**(23), 2423–2437 (2010)
- Davies, T.M., Jones, K., Hazelton, M.L.: Symmetric adaptive smoothing regimens for estimation of the spatial relative risk function. *Comput. Stat. Data Anal.* **101**, 12–28 (2016)
- Davies, T.M., Marshall, J.C., Hazelton, M.L.: Tutorial on kernel estimation of continuous spatial and spatiotemporal relative risk. *Stat. Med.* **37**(7), 1191–1221 (2018)
- Diggle, P.: A kernel method for smoothing point process data. *Appl. Stat.* **34**(2), 138–147 (1985)
- Diggle, P.: *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*, 3rd edn. CRC Press, Boca Raton (2014)
- Duyckaerts, C., Godefroy, G.: Voronoi tessellation to study the numerical density and the spatial distribution of neurones. *J. Chem. Neuroanat.* **20**(1), 83–92 (2000)
- Duyckaerts, C., Godefroy, G., Hauw, J.J.: Evaluation of neuronal numerical density by Dirichlet tessellation. *J. Neurosci. Methods* **51**(1), 47–69 (1994)
- Ebeling, H., Wiedenmann, G.: Detecting structure in two dimensions combining Voronoi tessellation and percolation. *Phys. Rev. E* **47**(1), 704–710 (1993)
- Ferreira, J., Denison, D., Holmes, C.: Partition modelling. In: Lawson, A., Denison, D. (eds.) *Spatial Cluster Modelling*, chap 7, pp. 125–146. CRC Press, Boca Raton (2002)
- Heikkinen, J., Arjas, E.: Non-parametric Bayesian estimation of a spatial Poisson intensity. *Scand. J. Stat.* **25**(3), 435–450 (1998)
- Holmström, L., Hamalainen, A.: The self-organizing reduced kernel density estimator. In: *IEEE International Conference on Neural Networks*, pp. 417–421 (1993)
- Kallenberg, O.: *Random Measures, Theory and Applications*. Springer, Berlin (2017)
- Last, G.: Stationary random measures on homogeneous spaces. *J. Theor. Probab.* **23**(2), 478–497 (2010)
- Lawrence, T., Baddeley, A., Milne, R.K., Nair, G.: Point pattern analysis on a region of a sphere. *Stat* **5**(1), 144–157 (2016)
- Levine, N.: Houston, Texas, metropolitan traffic safety planning program. *Transp. Res. Record J. Transp. Res. Board* **1969**, 92–100 (2006)
- Levine, N.: A motor vehicle safety planning support system: the Houston experience. In: Geertman, S., Stillwell, J. (eds.) *Planning Support Systems Best Practice and New Methods*, pp. 93–111. Springer, Dordrecht (2009)
- Loader, C.: *Local Regression and Likelihood*. Springer, New York (1999)
- McSwiggan, G., Baddeley, A., Nair, G.: Kernel density estimation on a linear network. *Scand. J. Stat.* **44**(2), 324–345 (2017)
- Møller, J., Rubak, E.: Functional summary statistics for point processes on the sphere with an application to determinantal point processes. *Spat. Stat.* **18**, 4–23 (2016)
- Møller, J., Schoenberg, F.: Thinning spatial point processes into Poisson processes. *Adv. Appl. Probab.* **42**(2), 347–358 (2010)
- Moradi, M.M., Rodríguez-Cortés, F.J., Mateu, J.: On the intensity estimator of spatial point patterns on linear networks. *J. Comput. Graph. Stat.* **27**(2), 302–311 (2018)
- Ogata, Y.: Significant improvements of the space-time etas model for forecasting of accurate baseline seismicity. *Earth Planets Space* **63**(3), 217–229 (2011)
- Okabe, A., Sugihara, K.: *Spatial Analysis Along Networks: Statistical and Computational Methods*. Wiley, Hoboken (2012)
- Okabe, A., Boots, B., Sugihara, K., Chiu, S.: *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, 2nd edn. Wiley, Hoboken (2000)
- Okabe, A., Satoh, T., Sugihara, K.: A kernel density estimation method for networks, its computational method and a gis-based tool. *Int. J. Geogr. Inf. Sci.* **23**(1), 7–32 (2009)
- Ord, J.: How many trees in a forest? *Math. Sci.* **3**, 23–33 (1978)
- Rakshit, S., Davies, T.M., Moradi, M.M., McSwiggan, G., Nair, G., Mateu, J., Baddeley, A.: Fast kernel smoothing of point patterns on a large network using 2D convolution. Submitted for publication (2018)
- Rakshit, S., Nair, G., Baddeley, A.: Second-order analysis of point patterns on a network using any distance metric. *Spat. Stat.* **22**, 129–154 (2017)
- Schaap, W.E.: DTFE: the Delaunay tessellation field estimator. PhD thesis, University of Groningen (2007)
- Schneider, R., Weil, W.: *Stochastic and Integral Geometry. Probability and Its Applications*. Springer, Dordrecht (2008)
- Scott, D.: *Multivariate Density Estimation. Theory, Practice and Visualization*. Wiley, New York (1992)
- Silverman, B.W.: *Density Estimation for Statistics and Data Analysis*. CRC Press, Boca Raton (1986)
- van Lieshout, M.N.M.: *Markov Point Processes and Their Applications*. Imperial College Press, London (2000)
- van Lieshout, M.N.M.: On estimation of the intensity function of a point process. *Methodol. Comput. Appl. Probab.* **14**, 567–578 (2012)
- Wand, M., Jones, M.: *Kernel Smoothing*. CRC Press, Boca Raton (1995)
- Xie, Z., Yan, J.: Kernel density estimation of traffic accidents in a network space. *Comput. Environ. Urb. Syst.* **32**(5), 396–406 (2008)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.